APPENDIX-VII

Limited Usefulness of Conventional Subject Headings and Classification Systems as Basis for Machine Searching

It would have been possible to use the new machines with any set of subject heading or any classification system that had been developed previously. This could have been done, to consider one possibility, by using subject headings as ordinarily set up. This procedure would not, however, have made broad generic terms, e.g. "insect", "weapon", available as reference points for guiding searches by machine, when the analysis of the subject matter led to specific terms, such as "mosquito" or "torpedo", as appropriate index entries. To provide both generic and specific terms for conducting machine searches and yet avoid unduly burdening the indexer, the semantic factoring technique, outlined in the preceding section of this report, was developed.

It would also have been possible tu use the new searching machines with any conventional classification system. A little reflection will make apparent the disadvantages of this approach. A conventional classification system operates by sorting out items, e.g., patents, into groups and storing them in pigeonholes or similar compartments. Once a large number of items have been incorporated into such a system it is an extremely time-consuming job to rearrange them to accommodate search requirements other than those that were foreseen at the time when the classification system was set up. Another way to make this basic point is to say that conventional classification, based on rigidly established compartments, provides static storage for information but that such storage by its very nature must fail to furnish dynamic means for accomplishing reordering or rearrangement to meet information requirements not visualized in advance.

The rigid framework of headings that constitute a conventional classification scheme are built up from basic concepts by establishing between them certain fixed relationships. These relationships are designed by the constructer of a conventional classification system to meet requirements that it is anticipated the system must serve. Such relationships are necessarily limited in number as otherwise the system becomes excessively complicated. This fact, coupled with the difficulty of predicting just what information requirements will arise, imposes severe restrictions on what can be accomplished with conventional classifications. The new searching techniques permit the relationships between concepts to be set up as search requirements arise and the number of such relationships is virtually

unlimited. Operating the machine so as to select on the basis of a newly established relationship between basic concepts is much easier if these have not been frozen into an array of the type that characterizes conventional classification. Fixed relationships, of the type involved in conventional classification, not only are unnecessary but also disadvantageous.

The program for developing new indexing methods for searching by electronic machines must involve a large measure of pioneering. Nevertheless, it was possible to conserve much time and effort by taking advantage of previously developed classification systems, subject heading lists, indexes of textbooks and the like.

We devoted particular attention to the Intelligence Subject Classification which had been developed by OCD in pioneering the use of standard IBM accounting machines for searching and correlating information. It was evident that we could have used our new machines with the ISC without changing or revising it. This would have permitted the ISC to attain a degree of effectiveness as a searching tool previously impossible with standard IBM equipment. Increased effectiveness would have resulted from the greater flexibility of the newly developed punching method which enables a series of subject headings to be punched on a single card and which permits a machine scanning operation to be directed regardless of their location on the card simultaneously to several subject headings. The ability to scan and select in this way would probably have improved the discriminating power of the ISC by at least an order of magnitude.

These limitations on the usefulness of the ISC with the new searching machines, did not blind us to the fact that ISC embodies a large amount of careful thought and practical experience in dealing with problems encountered by CIA. Specifically, the terminology from which the ISC is constructed has been selected and found effective for designating--in a rather general fashion--perhaps, but nevertheless effectively--those aspects of the subject matter of documents of importance to intelligence work. This means that the ISC can be used to very good advantage as the source of terminology for constructing the indexing and coding scheme for exploiting the new IBM equipment. Furthermore, the combinations of terms set up in the ISC when establishing the individual classification headings are not without importance. It is true that the nature of these configurations are such that they cannot be incorporated directly into the new indexing and searching system. However, the configurations of terms found in the ISC provide a basis for discerning some of the ways in which individual index entries in our new index and coding scheme will be set up in combinations when providing answers to search requests. What the new indexing and classifying scheme will be able to do is to go much further than has heretofore been possible in

providing a substantial infinity of possibilities for combining indexing terminology for the purpose of defining searches. While the existing ISC system is confined to some eight thousand pre-established combinations, classificatory headings and sub-headings, our new system will provide a substantial infinity of possible combinations as a basis for machine searching.